DIAGNOSIS OF SHWACHMAN-DIAMOND SYNDROME

Field of the Invention

The invention relates to methods for diagnosing and treating individuals with Shwachman-Diamond Syndrome and for detecting Shwachman-Diamond disease carriers. More specifically, the invention relates to the identification of the Shwachman-Bodian-Diamond Syndrome (*SBDS*) gene and the identification of mutations of this gene which are associated with Shwachman-Diamond Syndrome.

Background of the Invention

Shwachman-Diamond Syndrome (SDS [MIM 260400]) is an autosomal recessive disorder with clinical features including exocrine pancreatic insufficiency, haematological dysfunction, and skeletal abnormalities[1,2,3]. Patients with SDS have a high risk of bone marrow failure and are at risk of developing acute myelogenous leukaemia (AML). SDS is the second most common cause of pancreatic insufficiency after cystic fibrosis and involves the failure of development of the exocrine pancreas. Other manifestations include skeletal abnormalities and liver function abnormalities, the latter being notable in young patients.

Many SDS patients present with malabsorption and steatorrhea related to their pancreatic insufficiency. Many such children fail to thrive due to the malabsorption and also due to their disinclination to eat normally because of gastrointestinal upsets. The haematological dysfunction most consistently involves neutropenia but can also present as thrombocytopenia or pancytopenia. Serious consequences for SDS patients include recurring severe infections that can be life threatening if the diagnosis is not made with the provision of prompt treatments. Further, traditional methods for treatment of bone marrow failure are generally not successful in SDS patients at this time but the surveillance and monitoring of the bone marrow to determine the occurrence of myelodysplasia, aplastic anaemia and/or the development of AML do provide some options for intervention.

It is therefore important for the optimum development and overall long term prognosis of these children that they are diagnosed as having SDS as early as possible so that infections may be treated with appropriate interventions, so that blood and bone marrow can be monitored for cellularity (numbers and cell types) and so that pancreatic enzyme supplementation may be instituted to provide adequate or near normal food absorption.

There are other diseases associated with exocrine pancreatic dysfunction, such as Cystic Fibrosis and Pearson Marrow Syndrome, and other diseases such as congenital neutropenia, Blackfan-Diamond Syndrome and Fanconi Anaemia can mimic the haematological manifestations of SDS. It is important, for proper treatment, that SDS is diagnosed as early as possible but at present SDS can only be distinguished from other diseases causing similar symptoms by complex, symptom-based tests which may have to be repeated many times before a conclusion is reached (Rothbaum et al., (2002), J. Pediatrics, v. 141, pp. 266-270; Ginzberg et al., (2000), Am. J. Hum. Genet., v. 66, pp. 1413-1416).

There is therefore a real need for a convenient and definitive test, such as a genetic test or a gene product-based immunological test, to diagnose SDS. Further, as the bone marrow failure aspects are so serious, there is need to provide new options to correct the associated deficiencies. The identification and analysis of the gene that is affected in SDS would provide for such opportunities.

Segregation analysis of an international collection of families of SDS patients supports an autosomal recessive mode of inheritance (Ginzberg et al., (2000), Am. J. Hum. Genet., v. 66, pp. 1413-1416). Previous studies of families with SDS showed that the putative SDS locus mapped to the centromeric region of chromosome 7, to a 1.9 cM interval at 7q11[4,5]. The genetic defect associated with the disease has, however, not previously been identified.

Summary of the Invention

The invention provides a convenient and rapid method for the diagnosis of SDS, based on the finding that SDS is associated with mutations in a previously uncharacterised gene residing within the 1.9 centiMorgan disease interval at 7q11 delineated by linkage and haplotype analysis in family studies[4,5]. The gene, with a 1.6 kb transcript, was originally designated by the inventors as DEPCH and its encoded protein of 250 amino acids was designated depechin. The gene has been renamed as Shwachman-Bodian-Diamond Syndrome (*SBDS*) gene. A second copy previously designated DEPCHP and now designated *SBDSP*, with 97% nucleotide sequence identity, resides within a locally duplicated genomic block of at least 305 kb, and appears to be a pseudogene. Recurring mutations, the apparent result of recombination between the duplicated gene copies, were found in 89% of unrelated SDS patients (n=158), with 60% carrying two converted alleles and 29% having a different mutation in the second allele. The extent of the converted segments varied but consistently included at least one of two critical sequence changes predicted to result in truncation of the encoded protein. Other less common disease alleles involve missense and insertion/deletion changes distinct from those in the pseudogene. The gene is a member of a highly conserved protein family, with putative orthologues in diverse species ranging from archæbacteria to eukaryotes. The archæal orthologues are located within highly conserved operons that include homologues of genes involved in RNA processing[6], suggesting that SDS may be the result of a deficiency in some aspect of RNA metabolism that is essential for hæmatopoiesis, chondrogenesis and the development of the exocrine pancreas.

"*SBDS* or SBDS gene" is the chromosome 7q11.22 gene as described herein which when mutated is associated with SDS. This definition includes sequence polymorphisms wherein the nucleotide substitutions in the gene sequence do not affect the function of the gene product.

"SBDS protein" is the protein encoded by the *SBDS* gene.

"Mutant *SBDS* gene" is the *SBDS* gene containing one or more mutations which, if present on both alleles of the gene, lead to SDS.

In accordance with one embodiment, the invention provides a method for determining whether a subject is suffering from Schwachman-Diamond Syndrome (SDS) comprising

obtaining a nucleic acid sample from the subject, and

conducting an assay on the nucleic acid sample to determine the presence or absence of a *SBDS* gene mutation associated with SDS, wherein the presence of a *SBDS* gene mutation associated with SDS in both *SBDS* alleles indicates that the subject suffers from SDS.

In accordance with a further embodiment, the invention provides a method for determining whether a subject is an SDS carrier comprising

obtaining a nucleic acid sample from the subject, and

conducting an assay on the nucleic acid sample to determine the presence or absence of a *SBDS* gene mutation associated with SDS, wherein the presence of a *SBDS* gene mutation associated with SDS in one *SBDS* allele indicates that the subject is an SDS carrier.

In accordance with a further embodiment, the invention provides a method for determining whether a subject is suffering from Shwachman-Diamond Syndrome (SDS) comprising

obtaining a tissue sample from the subject, and

conducting an assay on the tissue sample to determine the level of SBDS protein in the sample, wherein a reduced level of SBDS protein in the sample relative to a control sample indicates that the subject suffers from SDS.

In accordance with a further embodiment, the invention provides a method for determining whether a subject is at risk for developing acute myelogenous leukaemia (AML) comprising

obtaining a nucleic acid sample from the subject, and

conducting an assay on the nucleic acid sample to determine the presence or absence of a *SBDS* gene mutation associated with SDS, wherein the presence of a *SBDS* gene mutation associated with SDS indicates that the subject is at risk for development of AML.

In accordance with a further embodiment, the invention provides a method for treating a subject suffering from SDS comprising administering to the subject a therapeutically effective amount of a substantially purified SBDS protein or of an isolated nucleotide sequence encoding an SBDS protein.

In accordance with a further embodiment, the invention provides an isolated nucleic acid molecule encoding an SBDS protein.

In accordance with a further embodiment, the invention provides an isolated nucleic acid molecule comprising at least about 10, 20, 30, 50, 75 or 100 consecutive nucleotides of SEQ ID NO:1 or 29.

In accordance with a further embodiment, the invention provides a substantially purified SBDS protein.

In accordance with a further embodiment, the invention provides an antibody which binds specifically to an epitope of an SDS protein.

In accordance with a further embodiment, the invention provides a nucleotide sequence selected from the group consisting of:

(a) 5'-GCGTAAAAAGCCACAATAC-3' (SEQ ID NO:3);
(b) 5'-CTATGACAGTATTCGTAAGACTAGG-3' (SEQ ID NO:4);
(c) 5'-GGGGATTTGTTGTGTCTTG-3' (SEQ ID NO:5);
(d) 5'-CTTTCCTCCAGAAAAACAGC-3' (SEQ ID NO:6);
(e) 5'-AAATGGTAAGGCAAATACGG-3' (SEQ ID NO:7);
(f) 5'-ACCAAGTTCTTTATTATTAGAAGTGAC-3' (SEQ ID NO:8);
(g) 5'-GCTCAAACCATTACTTACATATTGA-3' (SEQ ID NO:9);
(h) 5'-CACTTGCTTCCATGCAGA-3' (SEQ ID NO:10);
(i) 5'-AAAGGGTCATTTTAACACTTC-3' (SEQ ID NO:11);
(j) 5'-GAAAATATCTGACGTTTACAACA-3' (SEQ ID NO:12);
(k) 5'-TCCACTGTAGATGTGAACTAACTC-3' (SEQ ID NO:13);
(l) 5'-CACTCTGGACTTTGCATCTT-3' (SEQ ID NO:14);
(m) 5'-GCTTCTGCTCCACCTGAC-3' (SEQ ID NO:15);
(n) 5'AGCTATGCTGCAGCTGTTAC-3' (SEQ ID NO:16);
(o) 5'-ATGCATGTCCAAGTTTCAAG-3' (SEQ ID NO:17);
(p) 5'-TCCATGGCTATATTTTGATGA-3' (SEQ ID NO:18);
(q) 5'-TAAGCCTGCCAGACACAC-3' (SEQ ID NO:19);

(r) 5'-CACTCTGGACTTTGCATCTT-3' (SEQ ID NO:20);

(s) 5'-TGTTGGTTTTCACCGAATA-3' (SEQ ID NO:21);

(t) 5'-AGATAAAGAAAGACACACACAACT-3' (SEQ ID NO:22);

(u) 5'-GAAATCGCCTGCTACAAA-3' (SEQ ID NO:23);

(v) 5'-TCAGCTTCTTGCCTTCAT-3' (SEQ ID NO:24);

(w) 5'-TAAGTAAGCCTGCCAGACA-3' (SEQ ID NO:25);

(x) 5'-CATCAAGGTCTTTTTCCAAG-3' (SEQ ID NO:26);

(y) 5'-CCTGTCTCTGCCCAAGTC-3' (SEQ ID NO:27); and

(z) 5'-AGGGAACATTTTCAAAACTCA-3' (SEQ ID NO:28).

In accordance with a further embodiment, the invention provides a transgenic non-human mammal having within its genome an SBDS gene with at least one mutation associated with SDS.

In accordance with a further embodiment, the invention provides a kit comprising at least one pair of primers suitable for amplification of at least a portion of an SBDS gene.

Summary of Drawings

Fig. 1 shows an integrated map of the interval of chromosome 7 where the gene deficiency that leads to SDS resides. ***a***, The refined map interval, flanked by microsatellite markers D7S2429 and D7S502, is shown with reference to the Genbridge 3 radiation hybrid panel. ***b***, An expanded map of sub regions from RH bins 65 and 72 based on genomic sequences from BAC clones in GenBank. The regions contains at least 305 kb that has duplicated intrachromosomally. The positions and orientations of the paralogous duplicons along 7q were determined by unique STS content and radiation hybrid mapping. ***c***, Identified genes in the BAC contigs are shown. Duplicon A contains at least 2 genes, *SBDS* and *SDCR2A* (*Shwachman-Diamond Critical Region-2A*). ***d***, *SBDS* is composed of 5 exons (coding regions in grey, noncoding regions in black) spanning 7.9 kb of genomic sequence. The location of oligonucleotide primers used for mutation screening by genomic PCR and RT-PCR are indicated.

Fig. 2 shows mutations in *SBDS* associated with SDS. ***a***, Map of *SBDS* (coding regions in light blue) and sequence alignment of the exon 2 region of *SBDS* and *SBDSP*, with gene-specific sequences in green and pseudogene sequences in red. In comparison to *SBDS, SBDSP* exon 2 contains sequence changes (underlined in red) that are predicted to result in truncation of its predicted protein product. These include an in-frame stop codon at 184 bp and a T>C change at 250+10 bp (corresponding to the invariant T of the donor splice site at 258+2 bp in *SBDS*) which results in the use of an alternate donor splice site (invariant splice site positions are boxed) at 250+1 bp. The sequence differences in *SBDSP* present restriction sites for *Bsu36*I, and *Dde*I at 183 bp and *Cac8*I at 240+7 bp. ***b***, Electropherograms for cloned sequences from the exon 2 region of *SBDS* reveal sequence changes (red) derived from gene conversion events between *SBDS* and its pseudogene; three gene converted alleles are shown. These include *[183TA>CT]*, *[258+2T>C]*, and an extended conversion mutation *[183TA>CT +201A>G +258+2T>C]* with the intervening adenine (position 201) to guanine change. In each case, flanking sequences, including those at 129-2 bp and 258+124 bp, have not been converted (green). ***c***, A restriction map of the *SBDS* exon 2 amplimer (primers E and F, Fig. 1*d*) showing the position of *Cac8*I (C) and *Bsu36*I (B) restriction sites. Square brackets indicate the positions of restriction sites corresponding to converted sequences. The pedigree of family SW20 is shown with affected individuals in black and carriers in grey. Restriction fragment analysis of PCR amplified *SBDS* exon 2 sequences revealed that the brothers inherited *[183TA>CT]* through the father and paternal grandfather, and *[258+2T>C]* through the mother and maternal grandmother. Patient P1 is heterozygous for *[258+2T>C]* and the extended conversion mutation (*[183TA>CT +201A>G +258+2T>C]*). Two unrelated control individuals are also shown (C1 and C2). ***d***, Restriction maps of the gene and pseudogene loci showing the locations of all *Nde*I restriction sites (N). Hybridisation of a DNA probe derived from a partial *SBDS* cDNA (green) to genomic DNAs restriction digested with *Nde*I indicates that members of family SW6 (including patient P1 with two converted alleles) show a pattern of

hybridisation similar to two unrelated control individuals (C3 and C4) indicating that no rearrangements or deletions have occurred in the vicinity of *SBDS* or *SBDSP*. **e**, Sequence traces depicting other representative coding mutations in patient *SBDS* compared to controls (N), including an insertion (*[96_97insA]*), a deletion (*[119delG]*) and two missense mutations (*[24C>A]* and *[505C>T]*).

Fig. 3 shows expression analysis of *SBDS* and *SBDSP*. FTh Fetal thymus, FSp Fetal spleen, FLi Fetal Liver, FK Fetal kidney, FSM Fetal skeletal muscle, FLu Fetal lung, FH Fetal heart, FB Fetal brain, K Kidney, SM Skeletal muscle, Lu Lung, H Heart, B Brain, Li Liver, Pl Placenta, Pa Pancreas, Th Thymus, Sp Spleen, Ly Lymphocytes, To Tonsil, BM Bone Marrow, Le Peripheral Blood Leukocytes, LN Lymph Node, GAPDH Glyceraldehyde-3-Phosphate Dehydrogenase. ***a***, RNA expression survey of *SBDS* and *SBDSP* in primary tissues using a cloned RT-PCR product containing the entire *SBDS* open reading frame (primers T and R). Cumulative levels of both gene and pseudogene transcripts appear to be lower in thymus and bone marrow. An alternatively spliced product was detected in several tissues and was most prominent in peripheral blood leukocytes (Le). As shown in the lane indicated with an asterisk, this large transcript was detected with a probe derived from intron 1. ***b***, Analysis of patient EBV-transformed B lymphoblastoid-derived RNA shows that *SBDS* and *SBDSP* cumulative expression is lower in some patients compared to a control individual (C). The probe used to provide a control for RNA loading consisted of a 983bp cloned cDNA fragment from glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*). ***c***, RT-PCR expression analysis of *SBDS* and *SBDSP* was carried out with specific oligonuleotide primers and indicated that both transcripts are widely expressed. Sequencing of PCR products led to the identification of an exon 2^{minus} transcript. RT-PCR indicated that the alternatively spliced product (shown as 349bp) is present in all tissues tested, however its expression is significantly lower than transcripts that include exon 2 (shown as 479bp).

Fig. 4 shows CLUSTALX alignment of *SBDS*-encoded protein, SBDS, and representative orthologues. Strong conservation is seen throughout the

alignment from archæbacteria to complex eukaryotes. '*' represents absolutely conserved residues in the alignment, ':' represents positions at which conservative amino acid substitutions are observed and '.' represents semi conservative substitutions. The degree of sequence similarity is less pronounced towards the C-terminus although subgroups retain strong conservation. The human amino acid sequence (Hsa) is shown in bold. The locations of all identified coding mutations are represented as white letters on a black background and corresponding amino acid sequence changes are shown above the alignment. A putative U1-like zinc finger domain in three plant orthologues is indicated with a black bar. Ath *Arabidopsis thaliana*, Dme *Drosophila melanogaster*, Cel *Caenorhabditis elegans*, Mmu *Mus musculus*, Hsa *Homo sapiens*, Ola *Oryzias latipes*, Sce *Saccharomyces cerevisiae*, Ecu *Encephalitozoon cuniculi*, Mac *Methanosarcina acetivorans str. C2A*, Hnr *Halobacterium sp. NRC-1*, Mka *Methanopyrus kandleri str. AV19*, Mja *Methanococcus jannaschii*, Afu *Archaeoglobus fulgidus*, Pab *Pyrococcus abyssi*, Tac *Thermoplasma acidophilum*, Pae *Pyrobaculum aerophilum*, Sso *Sulfolobus solfataricus*, Ape *Aeropyrum pernix*, Pba *Populus balsamifera,* Gar *Gossypium arboreum*, [+]derived from partial GenBank EST sequence.

Fig. 5 shows the *SBDS* cDNA and its predicted encoded polypeptide. A: The nucleotide sequence of the cDNA corresponding to *SBDS* mRNA is shown numbered with the +1 starting at the first nucleotide, A, of the translation initiating codon. The 5' and 3' untranslated regions are shown in lower case, and the coding segment is shown in upper case text. B: amino acid sequence of the encoded polypeptide of 250 amino acids is shown numbered.

Fig. 6 shows the aligned genomic sequence for the human *SBDS* gene (SBDS) and its pseudogene *SBDSP* (SBDSP) and for the mouse *SBDS* gene (MUSBDS). The sequences for the five human exons are included with numbering that corresponds to that indicated in Fig. 5A. *SBDS* specific oligonucleotide primers that can be used to determine the nucleotide sequence of expressed RNA or of each of the exons for mutation detection are indicated by underlining of the SBDS sequence. Dual specific

oligonucleotide primers are indicated by the underlining of both *SBDS* and *SBDSP* sequences. The sequence of oligonucleotide primers indicated in the forward direction (the arrows pointing to the right) correspond directly to the sequence shown, while those primers in the reverse direction (the arrows pointing to the left) are comprised of the reverse complement of the indicated sequence.

Figure 7 shows the specificity and reactivity of antibodies produced to detect the SBDS protein. **a,** Polyclonal antibodies produced with recombinant SBDS (anti-rSBDS), left panel or a carboxyl peptide (anti-CpSBDS) of amino acids 224-239 (aa^{224}IKKETKGKGSLEVLNL^{239}) of SBDS, right panel, detected single bands of the predicted size in whole cell extracts of induced host *E. coli* BL21 containing the pET-28a expression vector with an in-frame fusion of the entire SBDS open reading frame. A polyclonal antibody to an amino peptide (anti-NpSBDS) of amino acids 32-47 (aa^{32}CYKNKVVGWRSGVEKD^{47}) of SBDS has also been generated, data not shown. **b**, The anti-rSBDS antibody also detected SBDS expressed transiently in HEK293 cells under the control of a CMV promoter. The bands corresponds to those detected by anti-Myc or anti-HA antibodies. The subtle shifts in sizes are due to the various epitope tags and/or their locations that have been fused in frame to the *SBDS* gene, including amino or carboxyl positioned Myc (N-Myc or C-Myc) N-HA or amino or carboxyl positioned HA (N-HA or C-HA) tags. **c**, Anti-rSBDS also detected a prominent band in whole cell extracts of the predicted size for SBDS in BxPC3 (ATCC CRL-1687), SV40-transformed human fibroblasts (GM00639), Caco-2 (ATCC HTB-37), AR42J (ATCC CRL-1492), EBV transformed human lymphoblast (GM003798), PANC1 (ATCC CRL-1469) and J.RT3 (ATCC TIB-153) cell lines. The total protein loaded per extract is as indicated below each panel.

Detailed Description of the Invention

The inventors have identified the *SBDS* gene and described the association of mutations in that gene with the autosomal recessive disease, SDS.

Clinical presentation in SDS can be variable but family studies have supported a single gene locus near the centromere at 7q11[2,4,5]. Eighteen positional candidate genes were identified in compiled genomic sequences from the locus, and eight of these were analysed for mutations in members of linked families. Disease-associated changes were identified in a gene represented by the full length, 1.6 kb cDNA clone flj10917 (OVARC1000321). The gene was initially designated by the inventors as *DEPCH* (*Development of Exocrine Pancreas, Chondrocytes and Hæmatological lineages*). The gene has been renamed (as approved by the Human Genome Organisation Gene Nomenclature Committee) as Shwachman-Bodian-Diamond Syndrome (*SBDS*) gene. The cDNA sequence is given in Fig. 5A (SEQ ID NO:1). *SBDS* is composed of 5 exons spanning 7.9 kb, and is contained in BAC clone RP11-325K1. The nucleotide sequences of the exons and surrounding introns are given in Fig. 6. The sequence of murine SBDS is also shown in Fig. 6. *SBDS* and part of an adjacent gene reside in a block of genomic sequence of at least 305 kb that is locally duplicated (Fig. 1). The paralogous duplicon was mapped distally, and contains an unprocessed pseudogene copy of *SBDS*, named *SBDSP*. The pseudogene transcript is 97% identical to the *SBDS* transcript with small deletions and single nucleotide changes that clearly disrupt coding potential.

The protein product encoded by *SBDS*, termed SBDS, is a member of a highly conserved protein family (Pfam UPF00023)[20]. Orthologues exist in species ranging from archæbacteria to vertebrates and plants (Fig. 4). The sequence of 250 amino acids is given in Fig. 5B (SEQ ID NO:2) for a predicted polypeptide of 28.8kDa with a pI of 8.9. The predicted amino acid sequence has no homology to any known functional domain, and no signal peptides were detected. The *S. cerevisiae* orthologue, encoded by ORF *YLR022c*, has been found to bind specifically and with high affinity to the phospholipids PI(4,5)P2 and PI(4)P using yeast proteome chips[21]. The gene has also been deleted by the Yeast ORF Deletion Project and haploid spores lacking *YRL022c* were found to be inviable[22]. Indirect lines of evidence suggest that orthologues of SBDS may play a role in RNA metabolism. First,

YLR022c has been clustered with other genes encoding RNA processing enzymes based on microarray expression profile analysis[23]. In addition, *SBDS* archæl orthologues are located in conserved operons that contain several RNA processing genes, including homologues of subunits of the eukaryotic exosome and RNaseP complexes[8]. The *A. thaliana* orthologue, along with sequences derived from partial cDNAs from *P. balsamifera* and *G. arboreum*, have extended carboxyl termini corresponding to putative RNA-binding domains, suggesting a functionally relevant fusion in flowering plants (Fig. 4). These observations suggest that SDS may be the result of a defect in an RNA processing pathway. Manifestation of disease must reflect the loss or perturbation of a cellular function that is particularly critical for the development of pancreatic acini, myeloid lineages, and chrondrocytes at growth plates of bones. The associated symptoms and the complications due to bone marrow failure may reflect not only the loss of one gene but also pleiotropic consequences of an aberrant pathway.

Sequence changes that do not alter protein-associated activities and that occur in normal individuals are likely to correspond to gene polymorphisms. A current accepted standard to discriminate polymorphisms from mutations is to screen 100 individuals of comparable ethnic background that are not affected with SDS. Examples of polymorphisms detected in *SBDS* are given in Table 2. SDS-associated mutations are shown in Table 1.

Diagnostic Methods

The invention provides a diagnostic method for determining whether a subject, such as a human subject, suffers from, or is at risk of developing, symptoms of SDS. In one embodiment, the method involves examining a nucleic acid sample from the subject for the presence or absence of a mutation of the *SBDS* gene associated with SDS. Such mutations include 183_184TA→CT; 183_184TA→CT+258+2T→C; 258+2T→C; 24C→A; 96-97insA; 119 delG; 131A→G; 199A→G; 258+1G→C; 260T→G; 291-293delTAAinsAGTTCAAGTATC; 377G→C; 505C→T; 56G→A; 93C→G; 97A→G; 101A→T; 123delC; 279_284delTCAACT; 296_299delAAGA;

354A→C; 428C→T+443A→G; 458A→G; 460-1G→A; 506G→C; and 624+1G→C. These mutations are identified in relation to the numbering of the nucleotide sequence of SEQ ID NO:1.

Many methods known to those of skill in the art can be used to detect the presence or absence of a *SBDS* gene mutation in the subject's nucleic acid.

The cDNA sequence of the wild type *SBDS* gene is shown in Figure 5 and is available at GenBank Accession Number AY169963 (NM_016038). The exon structure and flanking intron sequences are shown in Figure 6.

"Mutations" of the wild type *SBDS* gene associated with SDS include conversions, deletions, insertions, inversions or point mutations, either in the coding regions of the gene or gene regulatory regions.

A number of types of assay may be used to determine whether a subject has an *SBDS* gene mutation associated with SDS, including, for example, sequencing exons or other portions of the gene, including regulatory or intronic segments, PCR-RFLP analysis, allele specific PCR, allele specific oligonucleotide hybridisation restriction fragment length polymorphism (RFLP) analysis.

Where a direct sequencing assay is used, the sample may be DNA or RNA, for example genomic DNA or mRNA. Gene-controlling DNA segments and exons of an individual can be amplified and then examined for direct sequence changes, or scanned with methods that detect a heterozygous state followed by sequencing. These latter scanning methods can include single stranded conformational analysis (Orita M, Iwahana H, Kanazawa H, Hayashi K and Sekiya T (1989), "Detections of polymorphisms of human DNA by gel electrophoresis as single-stranded conformation polymorphisms", *Proc. Natl. Acad. Sci, USA* 86: 2776-2770), denaturing gradient gel electrophoresis (Wartell RM, Hosseini SH and Moran CP Jr (1990), "Detecting base pair substitutions in DNA fragments by temperature-gradient gel electrophoresis", (*Nucleic Acids Res.* 18: 2699-2705; Sheffield VC, Cox DR, Lerman LS and Myers RM (1989) or "Attachment of a 40-base-pair G + C rich sequence (GC clamp) to genomic DNA fragments by the polymerase chain reaction results in

improved detection of single-base changes" (*Proc. Natl. Acad. Sci, USA* 86: 232-236); and denaturing high pressure liquid chromatography Cotton RGH, Edkins E, Forrest S (eds) 1998 "Mutation detection: a Practical Approach" IRL Press, Oxford, and heteroduplex analysis Keen J, Lester D, Inglehearn C, Curtis A, Bhattacharya S (1991) Rapid detection of single base mismatches as heteroduplexes on Hydrolink gels. *Trends Genet.*, 7:5, amongst other methods. Larger deletions or insertions can be detected by traditional Southern blot analysis of DNA digest with restriction enzymes (Southern EM. (1975) 'Detection of specific sequences among DNA fragments separated by gel electrophoresis', J Mol Biol 98:503-17). Mutant alleles can be distinguished by observing their inheritance from each parent and although each patient will have two affected alleles, they will typically appear in heterozygous state (all of the references of this paragraph are incorporated herein by reference).

The diagnostic methods of the invention are used to screen subjects showing symptoms of possible SDS, such as pancreatic insufficiency to identify SDS, or to screen relatives of known SDS cases to determine whether they may be at risk of developing SDS symptoms.

The diagnostic method of the invention should preferably be carried out on samples from children at a young age in order to establish the diagnosis and allow appropriate treatment. The diagnostic method may also be used as a prenatal test, using amniotic fluid or CVS samples.

With respect to determining carrier status, as discussed below, the test may be carried out at any age, preferably at an age greater than 16 years in relatives of SDS patients.

Signs of SDS generally are evident in children at an early age and the diagnostic methods of the invention will usually be employed to determine if a child presenting with SDS symptoms is indeed suffering from SDS. On occasion, a sibling or close relative may be screened to determine if he or she suffers from SDS.

Suitable samples for testing of nucleic acid include buccal swabs, blood samples and bone marrow aspirates.

In one embodiment, genomic DNA is extracted from the sample and a target portion of the genomic DNA comprising the *SBDS* gene or a selected portion thereof is amplified by a polymerase chain reaction using suitable oligonucleotide primers, such as those described herein. The amplified nucleic acid is then sequenced using conventional techniques. The sequence is compared with the wild type sequence to determine the presence or absence of SDS-associated mutations. Primers must be selected which will amplify only the *SBDS* gene and not the pseudogene, as shown in Figure 6. Since a larger number of SDS-associated mutations have been observed in exon 2 of *SBDS* gene, it is preferable to look first for mutations in that exon. If no mutations are found in exon 2, exons 1 and 3 to 5 are similarly examined in turn.

One of skill in the art can select suitable primers by reference to the *SBDS* sequence of Figure 6, suitable primers are also identified in Example 1. Preferred primer pairs for amplification of *SBDS* exons are as follows:

Exon 1:	A & B or Q & B;
Exon 2:	E & F;
Exon 3:	G & H;
Exon 4:	SDCR9x4seqB;
	(5' - GCCTTCACTTTCTTCATAGT - 3') & J; and
Exon 5:	SDCR9x5Fseq
	(5' - GCTTGCCTCAAAGGAAGTT - 3') & L.

Regulatory regions of *SBDS*, such as the promoter region, may also be examined using suitable primers.

Promoter primers include SDCR9prom1RA (5' - CAGCCGACGACCTTGTTTT - 3') and SDCR9prom6FA (5' - GTGCCAACGCTGTGTTTT - 3').

These primers amplify a 501 bp segment partially overlapping exon 1, which likely contains the major controlling elements for the transcription of *SBDS* mRNA.

For conversion mutations found in exon 2, examination of the test subject's parents can be used to distinguish whether the subject has two

conversion recombinations rather than one extended conversion recombination.

In a further embodiment of the invention, an RNA sample is obtained from the test subject and is reverse transcribed by conventional methods to give a corresponding cDNA which is amplified by PCR and sequenced.

In a further embodiment, RFLP analysis may be used to detect *SBDS* gene mutations. Such methods of analysis are well known to those of skill in the art and an example is described in the Examples herein and in reference 30. Test samples are compared with normal controls and samples from patients with known mutations.

In a further embodiment, analysis of SBDS expression or of the level of SBDS protein may used to determine whether a subject suffers from or is at risk of SDS. As described herein, SBDS is expressed in a wide variety of tissues, including the most disease-relevant tissues, pancreas, bone marrow and myeloid cell lineages. A blood or tissue sample may therefore be used to evaluate SBDS expression or SBDS protein level. As seen in Figure 3b, mRNA level is notably reduced in SDS patients. SBDS expression can be evaluated by many routine methods, for example by mRNA analysis as described in the Examples herein and in reference 30.

In a further embodiment, an antibody specific for SBDS protein and carrying a detectable label can be used to assess the level of SBDS protein in a tissue sample of a subject by an immunological technique. Many suitable techniques, such as immunoprecipitation or ELISA assays, are known to those of skill in the art and are described, for example, in "Using antibodies - a laboratory manual", (1999), Harlow et al., Cold Spring Harbor Lab. Press. The level of protein in a test subject is compared with that in similar tissue samples from unaffected individuals, a reduction in level of SBDS protein being indicative of SDS. The identification of the *SBDS* gene and the absence of any known closely related homologues enables the preparation of antibodies highly specific for SBDS protein.

Detection of SDS Carriers

The invention further provides a method for determining whether a subject is an SDS carrier by determining whether the subject has an SDS-associated mutation in one allele of the *SBDS* gene.

The methods described above for detecting an SDS-associated mutation in a sample from a subject suspected of suffering from SDS may also be applied to detect carriers of the disease. The described methods for detecting such mutations in a nucleic acid sample from a subject are preferred.

Screening for SDS carriers is carried out especially on members of families with known SDS cases and may be important for genetic counselling of such family members regarding their likelihood of passing the disease on to their children. Generally, a method would be used to look for a specific mutation already found in an affected family member.

Identification of Further Mutations

The present invention also enables the identification of additional SDS-associated mutations of the *SBDS* gene, for example by examining SDS patients using the methods and primers described herein.

Amplification of target portions of the gene, followed by direct nucleic acid sequencing, as described herein for diagnostic purposes, and comparison with the wild type sequence, may be used to identify additional SDS-associated mutations.

Alternatively, assessment of the expression level of the *SBDS* gene, as described herein, may indicate reduced expression levels and point to further mutations which can be characterised by nucleic acid analysis as described above.

Nucleic Acids

The invention provides SBDS nucleic acids and homologues and portions thereof. Preferred nucleic acids have a nucleotide sequence which is at least 80%, preferably at least 90% and more preferably more than 97%

homologous to the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:29 or to a complement thereof.

Preferred nucleic acids are mammalian and especially preferred are human nucleic acids. Nucleic acids of the invention include nucleic acids encoding an amino acid sequence with at least 75%, preferably at least 90% and more preferably at least 99% amino acid identity to the amino acid sequence of SEQ ID NO:2, and nucleic acids encoding a portion of such amino acid sequences.

Also within the scope of the invention are nucleic acid molecules useful as probes or primers and comprising at least about 10, 20, 30, 50, 75, 90 or 100 consecutive nucleotides of SEQ ID NO:1.

Also within the scope of the invention are nucleic acids which hybridise under stringent conditions to a nucleic acid of the nucleotide sequence SEQ ID NO:1 or to a complement or a portion thereof. Stringent conditions for nucleic acid hybridisation are known to those skilled in the art and are described, for example, in "Protocols in Molecular Biology", (1989), John Wiley & Sons, N.Y., at 6.3.1 to 6.3.6.

Also within the scope of the invention are nucleic acids which differ from the sequence of SEQ ID NO:1 due to the degeneracy of the genetic code.

Proteins

The invention provides substantially purified SBDS proteins and portions thereof. These proteins and portions thereof are useful for the preparation of antibodies specific for SBDS proteins.

"Substantially purified" as used herein with respect to proteins means a protein preparation which is at least 75%, more preferably at least 90% and most preferably at least 99% by weight of SBDS protein.

Preferred SBDS proteins have an amino acid sequence which is at least about 75%, preferably at least about 90% and more preferably at least about 99% identical to the amino acid sequence of SEQ ID NO:2.

In a preferred embodiment, the SBDS protein has the amino acid sequence of SEQ ID NO:2. Full length proteins and portions thereof corresponding to one or more domains thereof or comprising at least 5, 10, 25, 50, 75 or 100 consecutive amino acids of SEQ ID NO:2 are within the scope of the invention.

The proteins and peptides of the invention may be isolated and purified by conventional protein purification methods such as gel-filtration chromatography, ion exchange chromatography, high performance liquid chromatography, immunoprecipitation or immunoaffinity purification.

SBDS proteins may be prepared by conventional recombinant methods, for example using the cDNAs described herein (for example human sequence has Genbank Accession Number AY169963) or a selected portion thereof. Since the *SBDS* gene is small, native gene expression may be achieved with the incorporation of natural promoter and enhancer gene elements. Suitable vectors and host cells for such expression are well known to those of skill in the art.

The expressed protein can be purified by standard procedures, as described above.

Antibodies

The present invention also enables the preparation of antibodies or antibody fragments which bind specifically to SBDS protein or to a portion thereof.

The term "antibody" means a monoclonal antibody or a polyclonal antibody, which binds specifically to a particular peptide, polypeptide or epitope, i.e. with greater affinity than to other peptides, polypeptides or eptiopes, and includes chimeric antibodies, humanised antibodies and single chain antibodies.

Chimeric antibodies are antibodies which contain portions of antibodies from different species. For example, a chimeric antibody may have a human constant region and a variable region from another species. Chimeric antibodies may be produced by well known recombinant methods, as

described in U.S. Patents Nos. 5,354,847 and 5,500,362, and in the scientific literature (Couto et al., (1993), Hybridoma, 12:485-489).

Humanised antibodies are antibodies in which only the complementarity determining regions, which are responsible for antigen binding and specificity, are from a non-human source, while substantially all of the remainder of the antibody molecule is human. Humanised antibodies and their preparation are also well known in the art – see, for example, U.S. Patents Nos. 5,225,539; 5,585,089; 5,693,761 and 5,693,762.

Single chain antibodies are polypeptide sequences that are capable of specifically binding a peptide or epitope, where the single chain antibody is derived from either the light or heavy chain of a monoclonal or polyclonal antibody. Single chain antibodies include polypeptides derived from humanised, chimeric or fully-human antibodies where the single chain antibody is derived from either the light or heavy chain thereof.

The term "antibody fragment" means a portion of an antibody that displays the specific binding of the parent antibody and includes Fab, F $(ab')_2$ and F_V fragments.

Polyclonal Antibodies

In order to prepare polyclonal antibodies, purified SBDS protein may be obtained, for example, as described herein. The purified protein or a portion thereof, coupled, if desired, to a carrier protein such as bovine serum albumin or keyhole limpet hemocyanin, as in Cruikshank WW, Center DM, Nisar N, Wu M, Natke B, Theodore AC, and Kornfeld H., (1994), Proc. Natl. Acad. Sci. USA 24: 5109-5113, is mixed with Fruend's adjuvant and injected into rabbits or other suitable laboratory animals.

Following booster injections at weekly intervals, the rabbits or other laboratory animals are then bled and the sera isolated. The sera can be used directly or purified prior to use by various methods including affinity chromatography employing Protein A-Sepharose, antigen Sepharose or Anti-mouse-Ig-Sepharose. Further purification methods well known in the art may be utilised to remove viral and/or endotoxin contaminants.

Monoclonal Antibodies

As will be understood by those skilled in the art, monoclonal antibodies may also be produced using an SBDS protein or a portion thereof. The protein or portion thereof, coupled to a carrier protein if desired, is injected in Freund's adjuvant into mice. After being injected three times over a three-week period, the mice spleens are removed and resuspended in phosphate buffered saline (PBS). The spleen cells serve as a source of lymphocytes, some of which are producing antibody of the appropriate specificity. These are then fused with a permanently growing myeloma partner cell, and the products of the fusion are plated into a number of tissue culture wells in the presence of a selective agent such as HAT. The wells are then screened by ELISA to identify those containing cells making binding antibody. These are then plated and after a period of growth, these wells are again screened to identify antibody-producing cells. Several cloning procedures are carried out until over 90% of the wells contain single clones which are positive for antibody production. From this procedure a stable line of clones which produce the antibody is established. The monoclonal antibody can then be purified by affinity chromatography using Protein A Sepharose, ion-exchange chromatography, as well as variations and combinations of these techniques. Truncated versions of monoclonal antibodies may also be produced by recombinant techniques in which plasmids are generated which express the desired monoclonal antibody fragment in a suitable host.

In a further embodiment, a cell line is provided which secretes an antibody specific for an SBDS protein or a portion thereof; a cell line secreting an antibody specific for a human SBDS protein is preferred.

Diagnosis of Predisposition to AML

A number of SDS patients have been found to develop AML. It is of some concern that individuals who have survived into adulthood without being diagnosed as SDS sufferers, because of minimal or unrecognised symptoms, may nevertheless also be at risk for the development of AML. The present

invention permits the identification of these individuals as SDS sufferers, so that they may be monitored for early signs of AML and appropriately treated. Although widespread screening of the population may not be practical, screening of relatives of diagnosed SDS patients for SDS-associated mutations is completely feasible, as also would be screening individuals exhibiting early or more overt signs of bone marrow transformation.

In addition, SDS carriers, who have an SDS-associated mutation in only one allele of the *SBDS* gene and are therefore asymptomatic, may be at risk for AML if they should experience loss or mutation of the wild-type allele, particularly in haemotological tissues. Again, screening of family members in SDS-affected families will indicate such genetic changes.

Kits

The invention further provides kits for use in the diagnostic methods described above for determining whether a subject is suffering from or is at risk for SDS, for determining whether a subject is a carrier of SDS or for determining whether a subject is at risk for AML. Such kits can comprise, for example, one or more pairs of oligonucleotide primers suitable for amplification of the SBDS gene or portions thereof, such as primers suitable for amplification of particular exons of SBDS, particularly human SBDS, as described for example in Figure 6. such kits can also contain instructions for use of the primers, and optionally, additional reagents required for the diagnostic methods described herein.

Therapeutic Methods

The invention further provides methods and compositions for treating subjects, including humans, suffering from SDS.

Methods of treatment are directed to restoring normal SBDS function in the subject.

Such methods include gene therapy to restore normal function at the gene level and administration of normal SBDS protein or portions thereof to make up for lack of normal gene expression.

Gene therapy may, for example, involve administration to the subject of a construct comprising an expression vector containing a nucleotide sequence encoding a wild type SBDS protein. Suitable expression vectors include retroviral, adenoviral and vaccinia virus vectors. Administration may be intravenous, oral, subcutaneous, intramuscular or intraperitoneal.

A large number of gene delivery methods are well known to those of skill in the art and may include, for example liposome-based gene delivery (Debs and Zhu (1993) WO 93/24640; Mannino and Gould-Fogerite (1988) BioTechniques 6(7): 682-691; Rose U.S. Pat No. 5,279,833; Brigham (1991) WO 91/06309; and Felgner et al. (1987) Proc. Natl. Acad Sci. USA 84: 7413-7414), and replication-defective retroviral vectors harboring a therapeutic polynucleotide sequence as part of the retroviral genome (see, e.g., Miller et al. (1990) Mol. Cell. Biol. 10:4239 (1990); Kolberg (1992) J. NIH Res. 4:43, and Cornetta et al. Hum. Gene Ther. 2:215 (1991)). Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof. See, e.g., Buchscher et al. (1992) J. Virol. 66(5) 2731-2739; Johann et al. (1992) J. Virol. 66 (5):1635-1640 (1992); Sommerfelt et al., (1990) Virol. 176:58-59; Wilson et al. (1989) J. Virol. 63:2374-2378; Miller et al.; J. Virol. 65:2220-2224 (1991); Wong-Staal et al., PCT/US94/05700, and Rosenburg and Fauci (1993) in Fundamental Immunology, Third Edition Paul (ed) Raven Press, Ltd., New York and the references therein, and Yu et al., Gene Therapy (1994) *supra*).

AAV-based vectors are also used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and in *in vivo* and *ex vivo* gene therapy procedures. See, West et al. (1987) Virology 160:38-47; Carter et al. (1989) U.S. Pat. No. 4,797,368; Carter et al. WO 93/24641 (1993); Kotin (1994) Human Gene Therapy 5:793-801; Muzyczka (1994) J. Clin. Invest. 94:1351 and Samulski (supra) for an overview of AAV

vectors. Construction of recombinant AAV vectors are described in a number of publications, including Lebkowski, U.S. Pat. No. 5,173,414; Tratschin et al. (1985) Mol. Cell. Biol. 5(11):3251-3260; Tratschin, et al. (1984) Mol. Cell. Biol. 4:2072-2081; Hermonat and Muzyczka (1984) Proc. Natl. Acad. Sci. USA 81:6466-6470; McLaughlin et al. (1988) and Samulski et al. (1989) J. Virol. 63:03822-3828. Cell lines that can be transformed by rAAV include those described in Lebkowski et al. (1988) Mol. Cell. Biol. 8: 3988-3996.

The organ with the most serious life threatening consequences, the bone marrow, may be treated by *ex vivo* gene therapy. This would involve the 1) extraction of bone marrow cells, 2) introduction of cDNA without mutations in conjunction with expression guiding elements followed by 3) re-introduction of these modified cells back to the bone marrow. Similar strategies have been used successfully in other diseases including severe combined immunodeficiency -X1 (M Cavazzana-Calvo, S Halcein-Bey, G de Saint Basile, F Gross, E Yvon, P Nusbaum, F Selz, C Hue, S Certain, J-L Casanova, P Bousso, F Le Deist and A Fischer. (2000) Gene therapy of human severe combined immunodefiency (SCID)-X1 disease. *Science* 288: 669-672; all of which are incorporated herein by reference). The *SBDS* gene is notably small such that native gene expression may be achieved with the incorporation of natural promoter and enhancer gene elements.

The *SBDS* nucleotide sequences described herein may be used in conventional expression systems, as described herein, to permit production of depechin protein in amounts sufficient for antibody production or for therapy.

Therapeutic compositions in accordance with the invention comprise an isolated nucleotide sequence encoding an SBDS protein or effective fragment thereof or a substantially purified SBDS protein or effective fragment thereof.

Transgenic animal models of SDS

The invention further enables the creation of an animal model of SDS which is important for further study of how SBDS mutations lead to the various SDS-associated disease manifestations and for testing of potential

therapeutics. A number of non-human mammals may be used to create such a model, including without limitation mice, rats, rabbits, sheep, goats and non-human primates. An animal model of SDS may have within its genome one or both *SBDS* genes with at least one mutation which when expressed results in symptoms of SDS. Identification and sequencing of the mouse *SBDS* gene homologue, as described herein, facilitates the creation of such animal models, for example a mouse model.

Methods for the creation of transgenic animals are well known to those of skill in the art. A transgenic animal according to the invention is an animal having cells that contain a transgene which was introduced into the animal or an ancestor of the animal at a prenatal (embryonic) stage. A transgenic animal can be created, for example, by introducing the gene of interest into the male pronucleus of a fertilised oocyte by, e.g., microinjection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The gene of interest may include appropriate promoter sequences, as well as intronic sequences and polyadenylation signal sequences. Methods for producing transgenic animals are disclosed in, e.g., U.S. Pat. Nos. 4,736,866 and 4,870,009 and Hogan et al., A Laboratory Manual, Cold Spring Harbor Laboratory, 1986. A transgenic founder animal can be used to breed additional animals carrying the transgene. A transgenic animal carrying one transgene can also be bred to another transgenic animal carrying a second transgene to create a "double transgenic" animal carrying two transgenes. Alternatively, two transgenes can be co-microinjected to produce a double transgenic animal. Animals carrying more than two transgenes are also possible. Furthermore, heterozygous transgenic animals, i.e., animals carrying one copy of a transgene, can be bred to a second animal heterozygous for the same transgene to produce homozygous animals carrying two copies of the transgene. For a review of techniques that can be used to generate and assess transgenic animals, skilled artisans can consult Gordon (Intl. Rev. Cytol., 115:171-229 (1989)), and may obtain additional guidance from, for example: Hogan et al, Manipulating the Mouse Embryo (Cold Spring Harbor Press, Cold Spring Harbor, N.Y. 1986); Krimpenfort et

al., Bio/Technology, 9:844-847 (1991); Palmiter et al., Cell, 41:343-345 (1985); Kraemer et al., Genetic Manipulation of the Early Mammalian Embryo (Cold Spring Harbor Press, Cold Spring Harbor, N.Y. 1985); Hammer et al., Nature, 315:680-683 (1985); Purscel et al., Science, 244:1281-1288 (1986); Wagner et al., U.S. Pat. No. 5,175,385; and Krimpenfort et al., U.S. Pat. No. 5,175,384.

EXAMPLES

The examples are described for the purposes of illustration and are not intended to limit the scope of the invention.

Methods of molecular biology, genetics, protein and peptide biochemistry and immunology referred to but not explicitly described in this disclosure and examples are reported in the scientific literature and are well known to those skilled in the art.

Methods

Human Subjects. Families with SDS included in this study have been described, and additional families have been obtained through ongoing recruitment[2]. The criterion for inclusion in the study was the presence of both exocrine pancreatic dysfunction and hæmatologic abnormalities, including neutropenia and other problems associated with bone marrow failure. Consent was obtained from all participating families, and procedural approval was obtained from the human subjects review board of The Hospital for Sick Children, Toronto (HSC). Genomic DNA was extracted either from Epstein-Barr virus (EBV) transformed B-lymphoblastoid cell lines or directly from peripheral white blood cell pellets, as described by Miller *et al.*[24]. Patient and control RNA was extracted from EBV-transformed B-lymphoblastoid cell lines as previously described[25]. DNA from 100 control Caucasian individuals (Human variation panel HD100CAU) was purchased from Coriell Cell Repositories (Camden, NJ).

Physical Mapping. Genomic sequences were identified through BLAST analysis of STSs and genetic markers in the SDS critical interval against the GenBank non-redundant (nr) and high throughput genome sequence (htgs) databases[26]. Where the density of pre-existing markers was low, BAC and YAC clones assigned to the region were subcloned and sequenced to provide new STSs as described[5]. Genomic sequences were compiled manually and the framework was supported by radiation hybrid mapping of select STSs.

Candidate Gene Identification. Candidate genes were identified in genomic sequences through the use of annotation data provided by GenBank (http://www.ncbi.nlm.nih.gov) and Project Ensembl (http://www.ensembl.org)[26,27]. *Ab initio* gene predictions were obtained through the use of GeneScript. Human genomic sequences were also compared to mouse genomic sequences (available through Celera Discovery System and Celera Genomics' associated databases) from the syntenic interval on mouse chromosome 5 using PipMaker2 to identify regions of cross-species conservation[28]. All *in silico* gene predictions were confirmed by RT-PCR analysis using random-primed cDNA derived from fetal brain, and/or testes poly(A)+ mRNA (Clontech, Palo Alto, CA).

Mutation Detection. The genomic structure of the *SBDS* gene and its pseudogene copy were used to design primer pairs using Primer3 to screen coding regions[29]. The position of primer pairs is shown (Figs. 1 and 6). PCR products were directly sequenced or cloned using a Topo TA-cloning kit (Clontech) prior to sequencing. Primer pairs (specific for *SBDS* unless otherwise stated) used were: A (5'-GCGTAAAAAGCCACAATAC-3') and B (5'-CTATGACAGTATTCGTAAGACTAGG-3') (exon 1), C (5'-GGGGATTTGTTGTGTCTTG-3') and D (5'-CTTTCCTCCAGAAAAACAGC-3') (exon 2, *SBDS/SBDSP* dual-specific), E (5'-AAATGGTAAGGCAAATACGG-3') and F (5'-ACCAAGTTCTTTATTATTAGAAGTGAC-3') (exon 2), G (5'-GCTCAAACCATTACTTACATATTGA-3') and H (5'-CACTTGCTTCCATGCAGA-3') (exon 3), I (5'-

AAAGGGTCATTTTAACACTTC-3') and J (5'-GAAAATATCTGACGTTTACAACA-3') (exon 4), K (5'-TCCACTGTAGATGTGAACTAACTC-3') and L (5'-CACTCTGGACTTTGCATCTT-3') (exon 5), M (5'-GCTTCTGCTCCACCTGAC-3') and N (5'AGCTATGCTGCAGCTGTTAC-3') (exons 1 & 2, *SBDS/SBDSP* dual-specific), O (5'-ATGCATGTCCAAGTTTCAAG-3') and P (5'-TCCATGGCTATATTTTGATGA-3') (exons 2 & 3, *SBDS/SBDSP* dual-specific). Patients were also screened for mutations through sequencing of RT-PCR products from random-primed cDNA derived from patient EBV-transformed B-lymphoblastoid cell lines. Primers used were: Q (5'-TAAGCCTGCCAGACACAC-3') and R (5'-CACTCTGGACTTTGCATCTT-3') (yields full length *SBDS* open reading frame), Q and S (5'-TGTTGGTTTTCACCGAATA-3'), and T (5'-AGATAAAGAAAGACACACACAACT-3') and R. Gene conversion mutations were detected through restriction analysis of exon 2 PCR fragments. Exon 2 was amplified from patient DNA using PCR primers C & D or E & F, and purified using a MinElute PCR Cleanup Kit (Qiagen). Restriction digestion using *Dde*I (not shown) or *Bsu036*I (*[183TA>CT]*) and *Cac8*I (*[258+2T>C]*) (New England Biolabs, Beverly, MA) was carried out as recommended by the manufacturer and analyzed by agarose gel electrophoresis. For all mutations, allele-specific oligonucleotide hybridisation to amplified *SBDS* exons from control individuals was carried out as described[30].

Southern Hybridisation. Genomic DNA from patients and control individuals was subjected to restriction digestion with *Nde*I (New England Biolabs) as recommended by the manufacturer and products were separated by agarose gel electrophoresis. The DNA was blotted and hybridised with a radiolabeled *SBDS* partial cDNA probe (exons 1-3) as described[30].

RT-PCR and RNA Blot Analysis. A panel of cDNAs derived from 22 adult and fetal tissues (Clontech) were analyzed by RT-PCR according the supplier's recommendations. Primers used were T and R (*SBDS*), and (5'-

TAAGTAAGCCTGCCAGACA-3') and (5'-CATCAAGGTCTTTTTCCAAG-3') (*SBDSP*). Primers used to assess the relative amount of *SBDS* exon 2 alternative splicing were U (5'-GAAATCGCCTGCTACAAA-3') and V (5'-TCAGCTTCTTGCCTTCAT-3'). RNA blots of poly(A)+ mRNA (Clontech) were hybridized to DNA probes labeled with [α^{32}P]-dCTP[30]. The *SBDS* probe was a cloned RT-PCR fragment (primers Q and R). The intron 1 probe was PCR amplified from genomic DNA using primers (5'-CCTGTCTCTGCCCAAGTC-3') and (5'-AGGGAACATTTTCAAAACTCA-3').

Sequence Alignment and Analysis. *SBDS* orthologues were identified through BLASTP analysis of amino acid sequences in the GenBank nr database, and through TBLASTN analysis of the GenBank EST database (dbEST). Sequences were aligned with CLUSTALX using default parameters followed by manual adjustment[31]. Amino acids were analysed for the presence of functional motifs using Pfam and associated databases (http://www.sanger.ac.uk/Software/Pfam/)[21].

Genbank Accession Numbers. *SBDS* consensus cDNA, AY169963 cDNA flj10917, AK001779; SDCR2A (cDNA flj10900), AK001762; SDCR3 (cDNA flj10099), AK000961; BAC RP11-458F8, AC073335; BAC RP11-325K1, AC079920; BAC RP11-584N20, AC069291; BAC RP11-324F21, AC073089; BAC RP11-166O4, AC006480; BAC RP11-479C13, AC005236. Depechin orthologues: *Arabidopsis thaliana At1g43860* gene product, NP_564488; *Drosophila melanogaster CG8549* gene product, NP_648057; *Caenorhabditis elegans* protein W06E11.4.p, NP_497226; *Mus musculus* protein 22A3, P70122; *Oryzias latipes* amino acid sequence derived from cDNA clone MF01SSA157A09 5' and 3' overlapping sequence reads, BJ013200 and BJ025159; *Saccharomyces cerevisiae* Ylr022cp, NP_013122; *Encephalitozoon cuniculi ECU08_1610* gene product, NP_597289; *Methanosarcina acetivorans str. C2A MA1778* gene product, NP_616704; *Halobacterium sp. NRC-1* Vng1276c, NP_280149; *Methanopyrus kandleri str. AV19 MK0384* gene product, NP_613669; *Methanococcus jannaschii MJ0592*

gene product, NP_247572; *Archaeoglobus fulgidus AF0491* gene product, NP_069327; *Pyrococcus abyssi PAB0418* gene product NP_126299; *Thermoplasma acidophilum Ta1291m* gene product, NP_394745; *Pyrobaculum aerophilum PAE2209* gene product, NP_559847; *Sulfolobus solfataricus SSO0737* gene product, NP_342243; *Aeropyrum pernix APE1167* gene product, NP_147753; *Populus balsamifera subsp. Trichocarpa* amino acid sequence derived from cDNA clone F038P45Y, BI121507; *Gossypium arboreum* amino acid sequence derived from cDNA clone GA__Ed0050B07f, BQ402534.

Example 1

RT-PCR analysis of several SDS patients with *SBDS*-specific oligonucleotide primers (indicated as RT-PCR primers Q and R in Fig. 1a and described in Fig. 6) revealed recurring sequence changes in exon 2, including a TA>CT dinucleotide change at position 183 or an 8 bp deletion at the end of the exon (the nucleotide numbering is described in Figs. 5 and 6). Analysis of *SBDS* genomic sequences confirmed the presence of the *[183TA>CT]* sequence change and revealed a *[258+2T>C]* nucleotide change in patients expressing the deleted *SBDS* transcript. *[258+2T>C]* is predicted to disrupt the donor splice site of intron 2, and the 8 bp deletion observed in the transcript is consistent with use of an upstream cryptic splice donor site at position 251. Alignment of patient *SBDS* sequences to genomic sequences from GenBank and control individuals indicated that both changes corresponded to sequences normally present in *SBDSP* (Fig. 2*a, b*). The dinucleotide alteration *[183TA>CT]* introduces an in-frame stop codon (K62X) while *[258+2T>C]* and its resultant 8 bp deletion also causes premature truncation of the encoded protein by frameshift (84Cfs3). Patient alleles were also identified that contain both of these changes together with an additional silent nucleotide change (*[201A>G]*) in the intervening segment, again consistent with the pseudogene sequence (Fig. 2*b*). The *[183TA>CT]* and *[258+2T>C]* changes could be detected in amplified *SBDS* genomic DNA followed by restriction digestion with *Bsu36*I and *Cac8*I, respectively (Fig. 2*a*,

c). Analysis of SDS pedigrees revealed that these changes were inherited and disease-associated. An example of segregating alleles in a linked pedigree is shown in Fig 2*c*. The specificity of genomic DNA amplimers for *SBDS* was supported by the absence of additional pseudogene-like sequence changes in nucleotide positions flanking the 183 and 258+2 bp positions (Fig. 2*b*) and the absence of any *SBDSP*-like sequences in 100 control samples. These findings, together with the observation of unaltered hybridisation patterns of genomic DNA with a *SBDS* probe (Fig. 2*d*), indicated that gene conversion due to recombination between *SBDS* and its highly homologous pseudogene had occurred. A similar basis for mutation has been observed in other genetic diseases[7-19]. Sequence analysis of the exon 2 region of patients indicated that most conversion events are confined to a short segment between 141 bp and 258+124 bp with a maximum size of 240 bp (Fig. 2*a*, *b*). Based on restriction digestion or sequencing of PCR products of patients from 158 unrelated families, 74% of SDS alleles (n=235 of 316) are the result of gene conversion, with 89% of patients carrying at least one converted allele and 60% carrying two converted alleles. Consistent with being a recessive disease, patients carry mutations on both copies of the *SBDS* gene. Of the patients analysed in the initial study, 50% were *[183TA>CT] + [258+2T>C]* compound heterozygotes, 5.1% were *[183TA>CT + 258+2T>C] + [258+2T>C]* compound heterozygotes, and 4.4% were homozygous for a *[258+2T>C]* conversion. Of patient alleles not displaying the conversion mutations, genomic sequencing revealed other changes within the coding region of *SBDS*, including small deletions, insertions, and nucleotide substitutions that would lead to frameshift and premature truncation, missense and nonsense changes (Table 1 and Fig. 4). To date, these mutations were not detected in 100 Caucasian control DNA samples by allele specific oligonucleotide hybridization or correspond to changes of highly conserved amino acids that would not be expected to be important for protein structure or function. Table 1 shows the SDS-associated mutations identified in the initial study and in subsequent studies.

Example 2

RNA hybridisation with *SBDS* indicated broad expression of a 1.6 kb message (Fig. 3*a*). Numerous GenBank EST clones, however, indicated that the pseudogene is also transcribed. Prominent larger-sized transcripts were also observed in poly(A)+ mRNA from several tissues and were confirmed to include intron 1 through hybridisation of an intron 1-specific probe (Fig. 3*a*). In addition, three GenBank EST clones corresponding to *SBDSP* were found to contain intron 1.

RNA expression analysis was carried out on a number of normal adult or fetal tissues, and on lymphoblasts from a number of SDS patients. As seen from Figure 3b, the level of combined *SBDS/SBDSP* mRNA, and consequently of protein product, was notably reduced in patient samples, compared with control C, lymphoblast RNA from a healthy subject.

Distinction between expression of the gene and pseudogene could be obtained through RT-PCR with specific oligonucleotide primers (Fig. 3*c*). Further, a broad survey of tissues revealed that the majority of *SBDS* mRNA does contain exon 2 although its alternative splicing was prominent in some patients (Fig. 3*c* and data not shown). Both RT-PCR and RNA analyses supported widespread expression of *SBDS* in all tissues examined, including the most disease-relevant tissues, pancreas, bone marrow, and myeloid lineages (Fig. 3*a*, *c*).

Example 3

Generation of antibodies for SBDS protein detection

Two methods were used to generate specific antibody probes to detect SBDS protein cells and tissues. First, a bacterially expressed polypeptide with the entire open reading frame of SBDS and, second, specified peptides synthesised from the amino and carboxyl portion (see legend to Fig. 7), were used as immunogens in rabbits. To obtain high level expression of recombinant SBDS, the complete open reading frame of the SBDS gene was incorporated into the pET28a vector (Novagen) using standard molecular biology techniques (Ref. 30). The open reading frame was fused with the

(HIS)6 tag of the expression vector for purification with immobilised metal (Ni2+) affinity chromatography. The purified polypeptide was then conjugated and injected into rabbits with the services of Washington Biotechnology, Inc. Pre-immune and immune sera were collected and whole cell protein extracts of various cell types were assessed, Fig. 7. The amino and carboxyl peptide antibodies were synthesised and prepared with the services of AnaSpec, Inc. and Washington Biotechnology, Inc., respectively. The antibodies showed high affinity and specificity for the SBDS protein product in different organs and cell lines, by Western blotting carried out as follows.

Whole cell extracts were prepared with Laemmli (*E. coli*) or RIPA (mammalian cells) buffer (and separated by 13.5% PAGE prior to blotting on Hybond C Extra (Amersham) membrane (Ref. 30 and Harlow and Lane). For rSBDS and anti-CpSBDS anti-sera, the membrane was blocked with 7% skim milk in TBST (10mM TrisHCl, pH7.3, 100mM NaCl with 0.1% Tween 20) for overnight at room temperature followed by incubation of a 1:2000 dilution for 5 h at room temperature. The blot was washed with TBST for five consecutive washes and incubated with anti-rabbit secondary antibody (Stressgen Biotechnologies Corp). The anti-Myc (Oncogene Research Products) and anti-HA (BAbCO-Covance) monoclonal antibodies and the anti-mouse secondary antibodies (Jackson ImmunoResearch Labs, Inc.) were used as recommended by their suppliers. The immunoreactive bands were detected by enhanced chemiluminescence.

Table 1: SDS-associated mutations

Nucleotide Sequence Changes	Predicted Amino Acid Change
183_184TA→CT	K62X
183_184TA→CT+258+2T→C	K62X
258+2T→C	84Cfs3
24C→A	N8K
96-97insA	N34fs15
119delG	S41fs17
131A→G	E44G
199A→G	K67E
258+1G→C	84Cfs3
260T→G	I87S
291-293delTAAinsAGTTCAAGTATC	D97-K98delinsEVQVS
377G→C	R126T
505C→T	R169C
56G→A	R19Q
93C→G	C31W
97A→G	K33E
101A→T	N34I
123delC	S41fs17
279_284delTCAACT	Q94_V95del
296_299delAAGA	E99fs20
354A→C	K118N
428C→T+443A→G	S143L + K148R
458A→G	Q153R
460-1G→A	splice
506G→C	R169P
624+1G→C	splice

Table 2: *SBDS Polymorphisms*

Some sequence changes in *SBDS* are predicted to be silent polymorphisms. Although some of these changes were detected in SDS patients, allele-specific oligonucleotide hybridisation was used to screen control samples to determine that these changes are not disease associated and should be classified as silent polymorphisms.

Nucleotide Sequence Change	Predicted Amino Acid Change
Intron 1	
129-71G→A	
129-185G→A	
129-225C→G	
129-265G→A	
Intron 2	
258+19A→G	
258+54T→G	
258+99A→C	
Intron 3	
459+92A→G	
Exon 2	
141C→T	L47L
201A→G	K67K
Exon 5	
651C→T	F217F
635T→C	I212T
Rare Change	
210T→C	D70E

The common mutations that account for the majority of SDS alleles can be detected by a PCR-restriction enzyme digestions of *Bsu*36I and *Cac*8I. These digestions can be performed singly (as described on page 27, methods) or in combination as detailed below. The combination digestion permits distinction between the conversions encompassing

PCR Amplification (same as for single digests already described)

Primer E (SEQ ID NO: 7): 5' –AAATGGTAAGGCAAATACGG- 3'
Primer F (SEQ ID NO: 8): 5' –ACCAAGTTCTTTATTATTAGAAGTGAC- 3'

product size: 733 bp
annealing temperature: 56.6 C
extension time: 40 sec

Double Digestion

*Bsu*36 I (New England Biolabs #R0524): 6 units plus
*Cac*8 I (New England Biolabs #R0579): 4.8 units
per 100-200ng PCR product
Digest 37 C >3hr

Band Sizes detected on agarose gel with ethidium bromide intercalation

Normal: 584bp also 64bp, 41bp, and smaller bands that are difficult to see
258+2 T>C: 431bp and 153bp also 64bp, 41bp and smaller bands
183 TA>CT: 358bp and 226bp also 64 bp, 41 bp and smaller bands
258+2T>C + 183TA>CT: 358bp, 153bp, 73bp also 64bp, 41bp and smaller bands

Note: Cannot use *Dde*I for this double digest, must use *Bsu*36I and *Cac*8I.

Mouse and human gene
88% nucleotide identity
97% amino acid identity

Dual Specific Digests for Common Mutations

PCR Amplification

Forward Primer: 5' –GGGGATTTGTTGTGTCTT- 3'
Reverse Primer: 5' – CTTTCCTCCAGAAAAACAGC - 3'

product size: 336 bp
annealling temperature: 56 °C
extension time: 1 min

Cac8 I Digest

Cac8 I (NEB #R0579): 4.8 units
Digest 37 °C >3hr

Band Size

Normal : 2 X 336 bp, 2 X 241 bp, 2 X 95 bp
1 allele with 258+2 T>C: 1 X 336 bp, 3 X 241 bp, 3 X 95 bp
2 alleles with 258+2 T>C: 4 X 241 bp, 4 X 95 bp

Dde I Digest

Dde I (NEB #R0175): 6 units
Digest 37 °C 2hr

Band Size

Normal : 2 X 190 bp, 2 X 169 bp, 4 X 146, 2 X 21 bp
1 allele with 183 TA>CT: 1 X 190 bp, 3 X 169 bp, 4 X 146, 2 X 21 bp
2 alleles with 183 TA>CT: 4 X 169 bp, 4 X 146, 2 X 21 bp

References

1. Shwachman, H, Diamond, L.K. & Khaw, K. The syndrome of pancreatic insufficiency and bone marrow dysfunction. *J. Pediatr.* **65**, 645-63 (1964).

2. Ginzberg, H. *et al.* Shwachman syndrome: Phenotypic manifestations of sibling sets and isolated cases in a large patient cohort are similar. *J. Pediatr.* **135**, 81-88 (1999).

3. Ginzberg, H. *et al.* Segregation analysis in Shwachman-Diamond syndrome: Evidence for Recessive Inhertiance. *Am. J. Hum. Genet.* **66**, 1413-1416 (2000).

4. Goobie, S. *et al.* Shwachman-Diamond syndrome with exocrine pancreatic dysfunction and bone marrow failure maps to the meric region of chromosome 7. *Am. J. Hum. Genet.* **68**, 1048-1054 (2001).

5. Popovic, M. *et al.* Fine mapping of the locus for Shwachman-Diamond syndrome at 7q11, identification of shared disease haplotypes, and exclusion of TPST1 as a candidate gene. *Eur. J. Hum. Genet.* **10**, 250-8 (2002).

6. Koonin, E.V., Wolf, Y.I. & Aravind, L. Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. *Genome Res.* **11**, 240-252 (2001).

7. Roesler, J. *et al.* Recombination events between the p47-phox gene and its highly homologous pseudogenes are the main cause of autosomal recessive chronic granulomatous disease. *Blood.* **15**, 2150-2156 (2000).

8. T. Strachan. Molecular pathology of 21-hydroxylase deficiency. *J. Inherit. Metab. Dis.* **17**, 430-41 (1994).

9. M.I. New. Steroid 21-hydroxylase deficiency (congenital adrenal hyperplasia). *Am. J. Med.* **98(1A)**, 2S-8S.

10. Beutler E. Gaucher disease as a paradigm of current issues regarding single gene mutations of humans. *Proc. Natl. Acad. Sci. USA.* **90,** 5384–5390 (1993).

11. Eikenboom, J.C., Vink, T., Briet, E., Sixma, J.J., and P.H. Reitsma. Multiple substitutions in the von Willebrand factor gene that mimic the pseudogene sequence. *Proc. Natl. Acad. Sci. USA.* **91**, 2221–2224 (1994).

12. Watnick, T.J., Gandolph, M.A., Weber, H., Neumann, H.P., and G.G. Germino. Gene conversion is a likely cause of mutation in PKD1. *Hum. Mol. Genet.* **7**, 1239–1243 (1998).

13. Chen, J.M., and C. Ferec. Molecular basis of hereditary pancreatitis. *Eur. J. Hum. Genet.* **8**, 473–479 (2000).

14. Chen, J.M., Raguenes, O., Ferec, C., Deprez, P.H., and C. Verellen-Dumoulin. A CGC>CAT gene conversion-like event resulting in the R122H mutation in the cationic trypsinogen gene and its implication in the genotyping of pancreatitis. *J. Med. Genet.* **37**, E36 (2000).

15. Cai, L. *et al.* A novel Q378Xmutation exists in the transmembrane transporter protein ABCC6 and its pseudogene: implications for mutation analysis in pseudoxanthoma elasticum. *J. Mol. Med.* **79**, 536-546 (2001).

16. Bunge, S., *et al.* Homologous nonallelic recombinations between the iduronate-sulfatase gene and pseudogene cause various intragenic deletions and inversions in patients with mucopolysaccharidosis type II. *Eur. J. Hum. Genet.* **6**, 492-500 (1998).

17. Hahnen, E., *et al.* Hybrid survival motor neuron genes in patients with autosomal recessive spinal muscular atrophy: new insights into molecular mechanisms responsible for the disease. *Am. J. Hum. Genet.* **59**, 1057-1065 (1996).

18. Campbell, L., Potter, A., Ignatius, J., Dubowitz, V., and K. Davies. Genomic variation and gene conversion in spinal muscular atrophy: Implications for disease process and clinical phenotype. *Am. J. Hum. Genet.* **61**, 40-50, (1997).

19. Wirth, B., et al. De novo rearrangements found in 2% of index patients with spinal muscular atrophy: Mutational mechanisms, parental origin, mutation rate, and implications for genetic counseling. *Am. J. Hum. Genet.* **61**, 1102-1111 (1997).

20. Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science.* **293**, 2101-2105 (2001).

21. A. Bateman, et al. The Pfam Protein Families Database. Nucl. Acids Res. 30(1):276-280 (2002).

22. Winzeler, E.A., *et al.* Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science.* **285**, 901-906 (1999).

23. Wu, L.F. *et al.* Large-scale prediction of Saccharomyces cerevisiae gene function using overlapping transcriptional clusters. *Nat. Genet.* **31**, 255-265 (2002).

24. Miller, S.A., Dykes, D.D., & H.F. Polesky. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**, 1215 (1988).

25. MacDonald, R.J., Smith, G.H., Przybyla, A.E., and J.M. Chirgwin. Isolation of RNA using guanidinium salts. *Meth. Enzymol.* **152**, 219-234 (1987).

26. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A., and D.L. Wheeler. GenBank. *Nucleic Acids Res.* **30**, 17-20 (2002).

27. T. Hubbard et al. The Ensembl genome database. *Nucleic Acids Res.* **30**, 38-41 (2002).

28. Schwartz et al. PipMaker - A Web Server for Aligning Two Genomic DNA Sequences *Genome Res.* **10**, 577-586 (2000).

29. Rozen, S. and H.J. Skaletsky. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, S., and S. Misener. Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ (2000).

30. Sambrook, J. and D.W. Russell. Molecular Cloning. Cold Spring Harbor Laboratory Press, NY (2001).

31. Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and D.G. Higgins. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**, 4876-4882 (1997).